

TMA4250: Project 2

Nora Røhnebæk Aasen and Elias Klakken Angelsen

May 5, 2022

Table of Contents

1 Problem 1	2
1.1 a)	2
1.2 b)	3
1.3 c)	4
1.4 d)	4
2 Problem 2	7
2.1 a)	8
2.2 b)	8
2.3 c)	10
2.4 d)	11
2.5 e)	14
2.6 f)	15

1 Problem 1

1.1 a)

The Besag model has density function

$$f(x; \kappa) \propto \kappa^{\frac{n-1}{2}} \exp\left(-\frac{\kappa}{2} \sum_{i \sim j} (x_i - x_j)^2\right),$$

for some precision parameter $\kappa > 0$. We wish to rewrite this in terms of a precision matrix \mathbf{Q} , such that

$$f(x; \kappa) \propto \kappa^{\frac{n-1}{2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^T \mathbf{Q}(\mathbf{x} - \mu)\right).$$

We can find a structure matrix \mathbf{R} , such that $\mathbf{Q} = \kappa \mathbf{R}$ and $\sum_{i \sim j} (x_i - x_j)^2 = \mathbf{x}^T \mathbf{R} \mathbf{x}$. This yields

$$f(x; \kappa) \propto \kappa^{\frac{n-1}{2}} \exp\left(-\frac{\kappa}{2} \sum_{i \sim j} (x_i - x_j)^2\right) = \kappa^{\frac{n-1}{2}} \exp\left(-\frac{1}{2} \mathbf{x}^T \kappa \mathbf{R} \mathbf{x}\right).$$

By expanding the square on the left side and pairing up the squares over the same i 's, we can choose R the following way, where n_i denotes the number of neighbors of node i . Define

$$R_{ij} = \begin{cases} n_i, & i = j \\ -1, & i \sim j, \\ 0, & \text{else.} \end{cases} \quad (1)$$

In our case, we are given the neighborhood matrices \mathbf{N}_1 and \mathbf{N}_2 describing which pairs of nodes (i, j) are related.

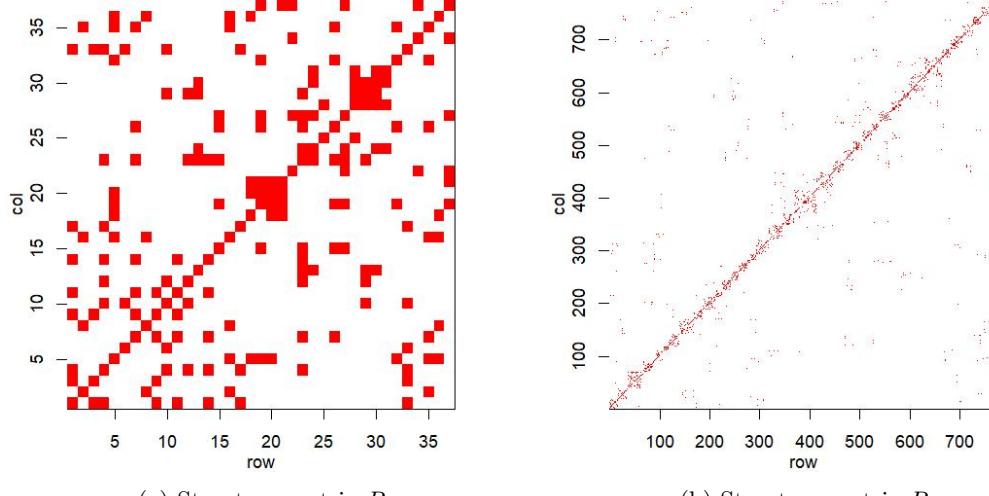
To find the structure matrices, we can consider $-\mathbf{N}_1$ and $-\mathbf{N}_2$ to get the off-diagonal entries of \mathbf{R}_1 and \mathbf{R}_2 . By adding the diagonal matrix containing the amount of neighbors for each node (found by summing each row in \mathbf{N}_k for $k = 1, 2$), we obtain R_k for $k = 1, 2$.

We find structure matrices \mathbf{R}_1 and \mathbf{R}_2 , and precision parameters τ_1 and τ_2 for the areas admin1 and admin2, respectively. The corresponding precision matrix for admin1 will therefore be $\mathbf{Q}_1 = \tau_1 \mathbf{R}_1$ and the precision matrix for admin2 is $\mathbf{Q}_2 = \tau_2 \mathbf{R}_2$.

The dimensions and ranks of the precision matrices are the same as the dimensions and ranks for the structure matrices. Therefore, we get dimension and rank 37×37 and 36 for admin1, and 775×775 and 774 for admin2. This is expected, as the dimension should tell us how many regions we are describing the neighborhood structure for. These matrices have rank one less than what would yield full rank, which is expected, as the Besag model is an intrinsic GMRF.

We notice from Figure 1a and Figure 1b that the sparsity pattern indicates a symmetric matrix, as we would expect.

The proportion of non-zero elements are 0.1527 for admin1 and 0.0088 for admin2, which is expected as there are a lot more regions in admin2, yielding more possibility for regions to not be neighboring.



(a) Structure matrix R_1 .

(b) Structure matrix R_2 .

Figure 1: Sparsity patterns for the two structure matrices used for each of our two admin areas.

We have to treat the Besag models as GMRFs as the precision matrix \mathbf{Q} is singular. For such a model, there exists no covariance matrix, so therefore it is not a standard multivariate Gaussian distribution. It is not even beneficial to perturb \mathbf{Q} to make it non-singular, as inverting the perturbed matrix yields dense matrices. Doing computations when assuming Besag models may work as standard multivariate Gaussian distributions is not efficient.

1.2 b)

To simulate from the Besag model on the admin1 graph with $\tau_1 = 1$ and sum-to-zero constraint, we simulate from a (first order) intrinsic GMRF using $\mathbf{Q}_1 = \mathbf{R}_1$. The simulation is done by sampling the proper part of the GMRF. By considering the kernel of \mathbf{Q}_1 to be $\text{span}\{(1, 1, \dots, 1)^T\}$, we invoke a sum-to-zero constraint on \mathbf{x} to shift the mean to 0 and therefore model the distance from the mean. The proper part we want to simulate is $\mathbf{x} | \sum x_i = 0$.

Algorithm 1 - How to simulate from first order intrinsic GMRFs

Input:

Q : a rank $n - 1$ semi-positive definite matrix with kernel = $\text{span}\{(1, 1, \dots, 1)^T\}$

ε : $0 < \varepsilon << 1$

Algorithm:

1. Define $\tilde{Q} = Q + \varepsilon I_n$
2. Compute \tilde{L} , the Cholesky factor of \tilde{Q}
3. Sample $z \sim \mathcal{N}_n(0, I_n)$
4. Solve the problem $\tilde{L}^T v = z$
5. Compute $\mathbf{x} = v - \text{mean}(v)(1, \dots, 1)^T$

Return:

x : The value simulated from the proper part of the GMRF, $\mathbf{x} | \sum x_i = 0$.

To simulate the Besag model on the admin1 graph with $\tau_1 = 1$, we input the matrix $\mathbf{Q}_1 = \mathbf{R}_1$

into Algorithm 1 and choose a small ε . To simulate from $\mathcal{N}_{37}(0, I_{37})$, we simulate 37 independent standard normal variables and list them as a vector. Therefore, we should expect the latter model to express no dependence between the regions due to the fact that Q_{ij} is nonzero only for $i = j$, while the Besag model (with sum-to-zero constraint) should give regional dependence based on the structure of the structure matrix indicating neighboring regions.

The sum-to-zero constraint projects the improper GMRF to the proper part, yielding a computationally convenient distribution to sample from. By requiring $\sum x_i = 0$, we invoke $\bar{\mathbf{x}} = 0$, as well as having linear dependence. Therefore, the freedom we have in choosing one of the x_i 's from the linear dependence makes the true mean unknown, suggesting that we are modelling the intrinsic distances from the mean, i.e. $\mathbf{z} = \mathbf{x} - \bar{\mathbf{x}}$. By knowing the precision matrix, we can therefore say something about how the data varies regionally, without knowing the scales at which the data operates (or even what the data describes).

We simulated values for our admin1 area using both a Besag model and a multivariate standard normal. The results can be seen in Figure 2. From the maps it seem to be regional correlation in the simulations using the Besag model, but not in the other model. This is as expected. Furthermore, the values generated from the Besag model is random around 0, since we have no knowledge of the mean in each area at this point. There is also no consistency in the values for each area between subfigure 2a and 2b, and 2c and 2d. Since both models are normally distributed with mean zero, we see that both simulations take values around 0. However, the correlation constrains the Besag simulations such that these values stay closer to 0.

1.3 c)

We simulated values for our admin2 area using both a Besag model and a multivariate standard normal. The results can be seen in Figure 3. We see the same tendencies as we discussed in section 1.2. Furthermore, as we have more regions in the admin2 graph than for admin1, it is easier to observe the differences. For the second parts of Figure 3, we randomly have larger and smaller values side-by-side, but for the Besag model, it is clear that the regional correlation is present, even though there are a few swift changes from large to small values.

Similar to what we observed in Figure 2, the regions in the Besag model are held back by their neighbors, causing less variation in the values they take, and thus yielding a lower span of values. This tendency is also present when we have more regions, and it is also clear that this is not a coincidence when we have as many regions as in admin2.

1.4 d)

In Figure 4 the empirical marginal variance for each state, calculated from 100 realizations from the Besag model, has been plotted. Based on this plot, we would say that our data does not appear to be stationary, since the variance is clearly not constant.

Next, we consider state 150, Gubio, and compute the correlation between this state and each other state. The resulting plot can be seen in Figure 5. Gubio is located in the upper right corner, and we can see that it is dependent on states all across the country.

The Besag model satisfies the pairwise Markov property, i.e.

$$X_i \perp X_j | \mathbf{X}_{-\{i,j\}}, \quad \{i, j\} \notin \mathcal{E}, i \neq j,$$

and since our data is drawn from a Besag model, our data should also satisfy this property. However, the pairwise Markov property only says that two non-neighbouring states will be

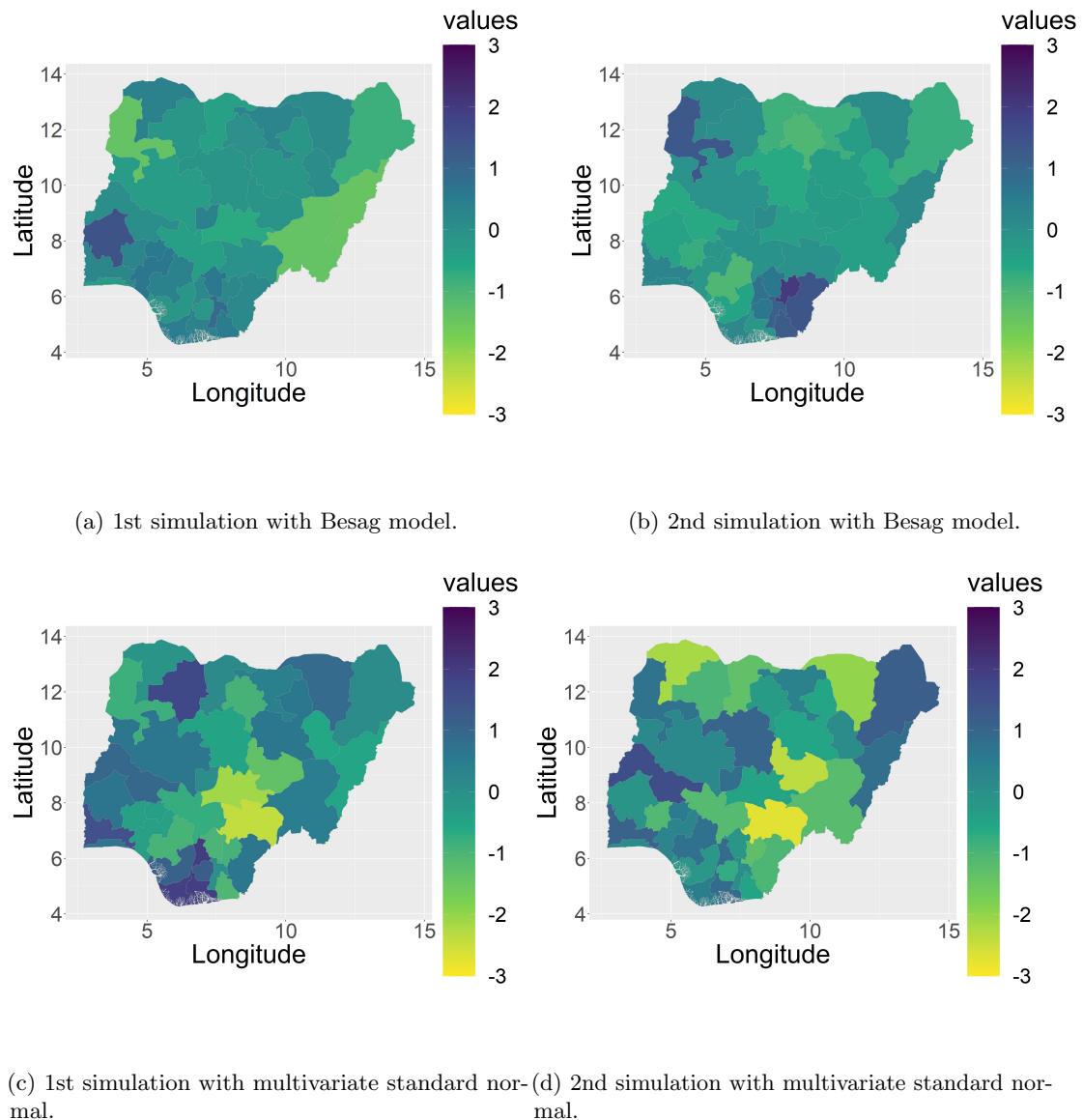


Figure 2: Plot of simulations for the admin1 areas.

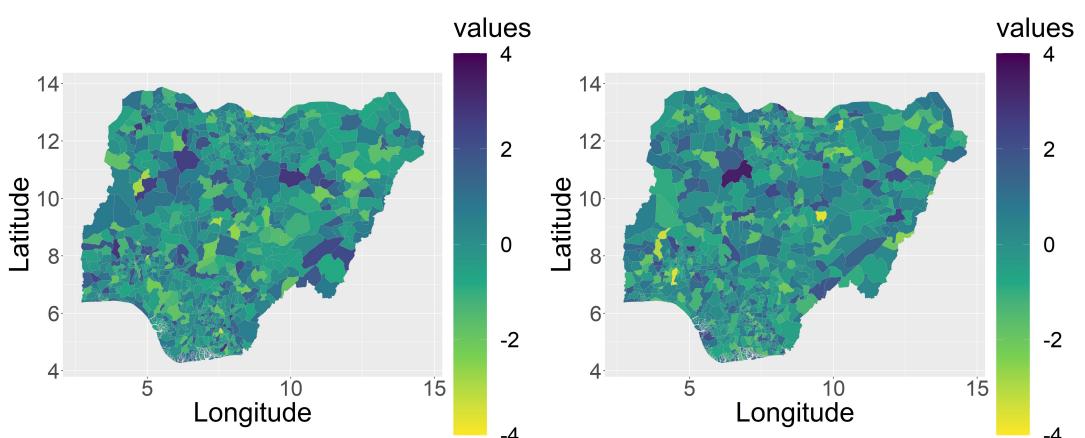
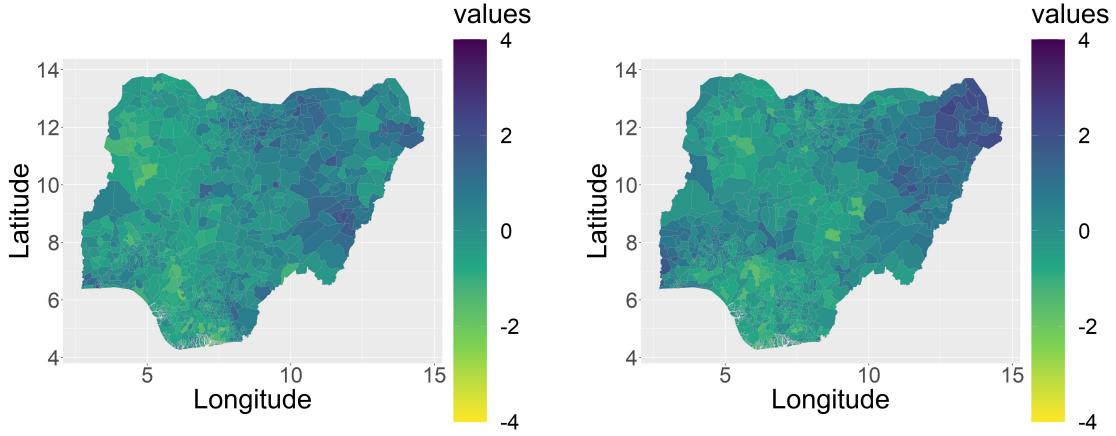


Figure 3: Plot of simulations for the admin2 areas.

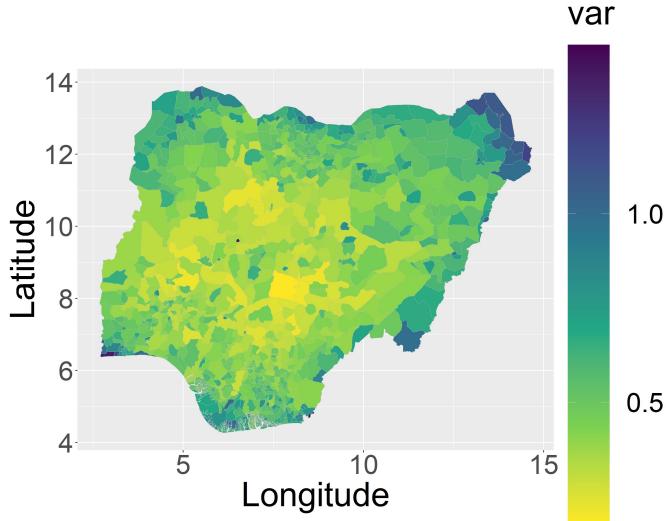


Figure 4: The marginal variance in each state.

independent *conditionally* on all neighbouring states. Hence, in our case we have not conditioned on the neighbours and since most states depends on its neighbours, there will be some flow of information across the entire country, although states further away has significantly lower correlation than those who are closer.

Without any assumption or constraint, we would not expect there to be negative correlations in a Besag model, but due to the sum-to-zero-constraint, we get negative correlation. This constraint assumes that

$$\sum x_i = 0,$$

and thus, as long as we have positive values, there must exist negative values as well. This is what causes the negative correlation in our model. However, we notice that the negative correlation is furthest away from Gubio, thus indicating that we do not expect negative correlation between neighbors in a Besag model.

2 Problem 2

In this problem we consider the vaccination coverage in the 37 areas in admin1. We will use hierarchical spatial models, where we assume that $\mathbf{X} = (\text{logit}(P_1), \dots, \text{logit}(P_{37}))$ is a vector containing the logit of the true proportions of vaccinated children, and $\mathbf{Y} = (\text{logit}(\hat{P}_1), \dots, \text{logit}(\hat{P}_{37}))$ is a vector containing the logit of the observed proportions of vaccinated children. We assume the relation

$$y_i \sim \mathcal{N}(x_i, V_i), \quad i = 1, \dots, 37$$

where V_i is some known variance.

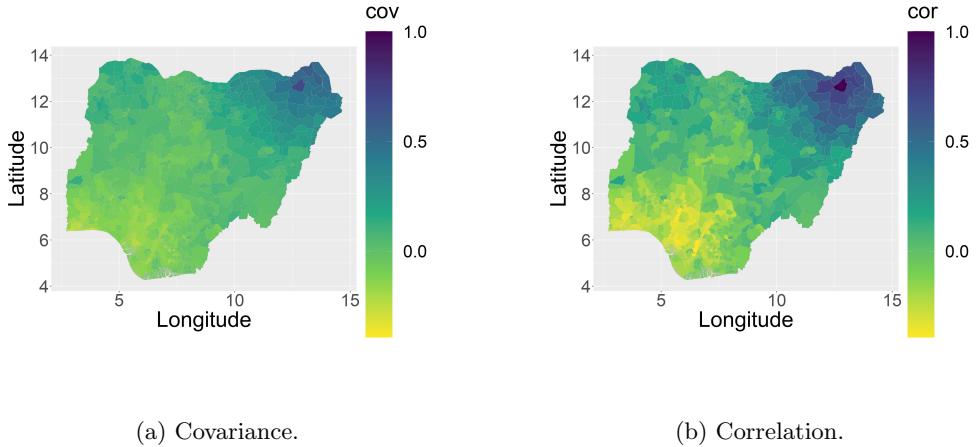


Figure 5: Covariance and correlation between each state and state 150, Gubio.

2.1 a)

From Figure 6, we see what appear to be spatial correlation in vaccination coverage. In particular, the observed proportions is low for the upper left region, and high for the lower, middle section of the map. Although a spatial model would introduce some bias, the assumptions are not unreasonable, and therefore it is a good measure for reducing variance.

2.2 b)

The distribution of $\mathbf{Y}|\mathbf{X}$ is the joint distribution of $y_i|x_i$ for all $i = 1, \dots, 37$. This is

$$\begin{aligned} p(\mathbf{Y}|\mathbf{X}) &= \prod_{i=1}^{37} f(y_i|x_i) \\ &= \prod_{i=1}^{37} \frac{1}{\sqrt{2\pi V_i}} \exp\left(-\frac{1}{2} \frac{(y_i - x_i)^2}{V_i}\right) \\ &= (2\pi)^{-\frac{37}{2}} \left(\prod_{i=1}^{37} V_i\right)^{-1/2} \exp\left(-\frac{1}{2} \sum_{i=1}^{37} \frac{(y_i - x_i)^2}{V_i}\right). \end{aligned}$$

This is a multivariate normal distribution with parameters

$$\mu = \mathbf{x}, \quad \Sigma = \begin{bmatrix} V_1 & 0 & \dots & 0 \\ 0 & V_2 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & V_{37} \end{bmatrix}.$$

From Bayes' theorem, we know that $p(\mathbf{X}|\mathbf{Y}) \propto p(\mathbf{Y}|\mathbf{X})p(\mathbf{X})$. If we assume a priori that $\mathbf{X} \sim \mathcal{N}_{37}(0, \sigma^2 \mathbf{I})$, then

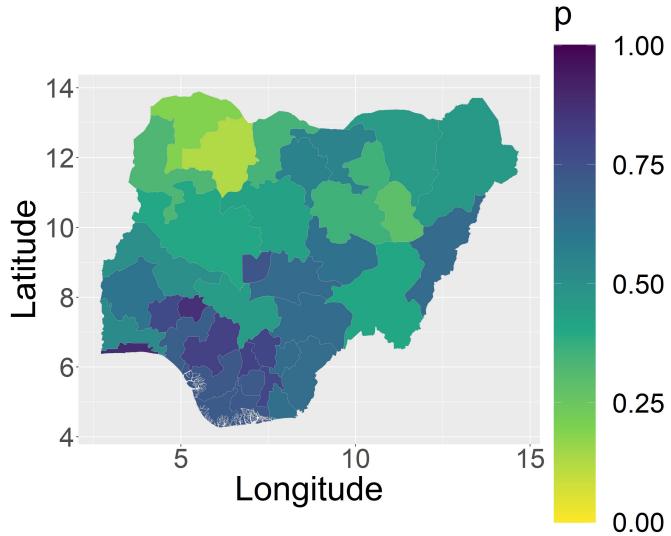


Figure 6: Observed proportions \hat{p} for each admin1 area.

$$\begin{aligned}
p(\mathbf{X}|\mathbf{Y}) &\propto (2\pi)^{-\frac{37}{2}} \left(\prod_{i=1}^{37} V_i \right)^{-1/2} \exp \left(-\frac{1}{2} (\mathbf{y} - \mathbf{x})^T \Sigma^{-1} (\mathbf{y} - \mathbf{x}) \right) \\
&\times (2\pi\sigma^2)^{-\frac{37}{2}} \exp \left(-\frac{\mathbf{x}^2}{2\sigma^2} \right) \\
&\propto \exp \left(-\frac{1}{2} \left((\mathbf{y} - \mathbf{x})^T \Sigma^{-1} (\mathbf{y} - \mathbf{x}) + \frac{\mathbf{x}^2}{\sigma^2} \right) \right) \\
&= \exp \left(-\frac{1}{2} \sum_{i=1}^{37} \left(\frac{(y_i - x_i)^2}{V_i} + \frac{x_i^2}{\sigma^2} \right) \right) \\
&= \exp \left(-\frac{1}{2} \sum_{i=1}^{37} \left(\frac{\sigma^2 y_i^2 - 2\sigma^2 y_i x_i + \sigma^2 x_i^2 + V_i x_i^2}{V_i \sigma^2} \right) \right) = (*)
\end{aligned}$$

We need to complete the square, and get this on the form of a multivariate normal distribution.

$$\begin{aligned}
(*) &= \exp \left(-\frac{1}{2} \sum_{i=1}^{37} \left(\frac{(\sigma^2 + V_i)x_i^2 - 2\sigma^2 y_i x_i + \frac{(\sigma^2)^2}{\sigma^2 + V_i} y_i^2 - \frac{(\sigma^2)^2}{\sigma^2 + V_i} y_i^2 + \sigma^2 y_i^2}{V_i \sigma^2} \right) \right) \\
&\propto \exp \left(-\frac{1}{2} \sum_{i=1}^{37} \left((\sigma^2 + V_i) \frac{x_i^2 - \frac{\sigma^2}{\sigma^2 + V_i} \sigma^2 y_i x_i + \frac{(\sigma^2)^2}{(\sigma^2 + V_i)^2} y_i^2}{V_i \sigma^2} \right) \right) \\
&= \exp \left(-\frac{1}{2} \sum_{i=1}^{37} \left(\frac{(x_i - \frac{\sigma^2}{\sigma^2 + V_i} y_i)^2}{\frac{V_i \sigma^2}{\sigma^2 + V_i}} \right) \right)
\end{aligned}$$

Hence, if we assume $\sigma^2 = 100^2$ we get that $p(\mathbf{X}|\mathbf{Y})$ is a multivariate normal distribution,

$$\mathbf{X}|\mathbf{Y} \sim \mathcal{N}_{37} \left(\left(\frac{1}{100^2} \mathbf{I} + \Sigma^{-1} \right)^{-1} (\Sigma^{-1} \mathbf{y}), \left(\frac{1}{100^2} \mathbf{I} + \Sigma^{-1} \right)^{-1} \right),$$

where Σ is the same covariance matrix as we defined earlier.

If we now let $\sigma^2 \rightarrow \infty$, and study the expression in our sum, we see that we get

$$\lim_{\sigma^2 \rightarrow \infty} \exp \left(-\frac{1}{2} \sum_{i=1}^{37} \left(\frac{(x_i - \frac{\sigma^2}{\sigma^2 + V_i} y_i)^2}{\frac{V_i \sigma^2}{\sigma^2 + V_i}} \right) \right) = \exp \left(-\frac{1}{2} \sum_{i=1}^{37} \left(\frac{(x_i - y_i)^2}{V_i} \right) \right).$$

Thus, when $\sigma^2 \rightarrow \infty$ we have $\mathbf{X} \sim \mathcal{N}_{37}(\mathbf{y}, \Sigma)$. Hence, $P_i|\mathbf{Y} = \mathbf{y} \sim \text{expit}(\mathcal{N}(y_i, V_i))$.

We use the conditional distribution $p(\mathbf{X}|\mathbf{Y})$ with $\sigma^2 = 100^2$ and simulate 100 samples for each region. From these samples we calculate the median and the coefficient of variation, defined as $c_v = \frac{\sigma}{\mu}$. These plots can be seen in Figure 7.

If we compare the estimated medians with what we saw in Figure 6, we see that the plots are almost identical. This makes sense, since we assumed no correlation between regions and very high variance in our prior. This results in low bias, and the observed proportions have the most influence on the conditionally simulated values.

2.3 c)

In this section we will repeat section 2.2, but instead of assuming a normal distribution as a prior for \mathbf{X} , we assume a prior that \mathbf{X} follows a Besag model with precision parameter τ and structure matrix \mathbf{R} , as defined in equation (1). We still assume that $p(\mathbf{Y}|\mathbf{X})$ follows a normal distribution, with expectation \mathbf{X} and precision matrix $\mathbf{Q} = \Sigma^{-1}$. We wish to determine the expectation and precision matrix of $\mathbf{X}|\mathbf{Y}$. First, we compute $p(\mathbf{X}, \mathbf{Y}) = p(\mathbf{X})p(\mathbf{Y}|\mathbf{X})$.

$$\begin{aligned}
p(\mathbf{X}, \mathbf{Y}) &\propto \exp \left(-\frac{1}{2} (\mathbf{y} - \mathbf{x})^T \mathbf{Q} (\mathbf{y} - \mathbf{x}) \right) \times \exp \left(-\frac{1}{2} \mathbf{x}^T \tau \mathbf{R} \mathbf{x} \right) \\
&= \exp \left(-\frac{1}{2} \mathbf{x}^T \tau \mathbf{R} \mathbf{x} - \frac{1}{2} \mathbf{y}^T \mathbf{Q} \mathbf{y} + \frac{1}{2} \mathbf{x}^T \mathbf{Q} \mathbf{y} + \frac{1}{2} \mathbf{y}^T \mathbf{Q} \mathbf{x} - \frac{1}{2} \mathbf{x}^T \mathbf{Q} \mathbf{x} \right) \\
&= \exp \left(-\frac{1}{2} \mathbf{x}^T (\tau \mathbf{R} + \mathbf{Q}) \mathbf{x} + \frac{1}{2} \mathbf{x}^T \mathbf{Q} \mathbf{y} + \frac{1}{2} \mathbf{y}^T \mathbf{Q} \mathbf{x} - \frac{1}{2} \mathbf{y}^T \mathbf{Q} \mathbf{y} \right)
\end{aligned}$$

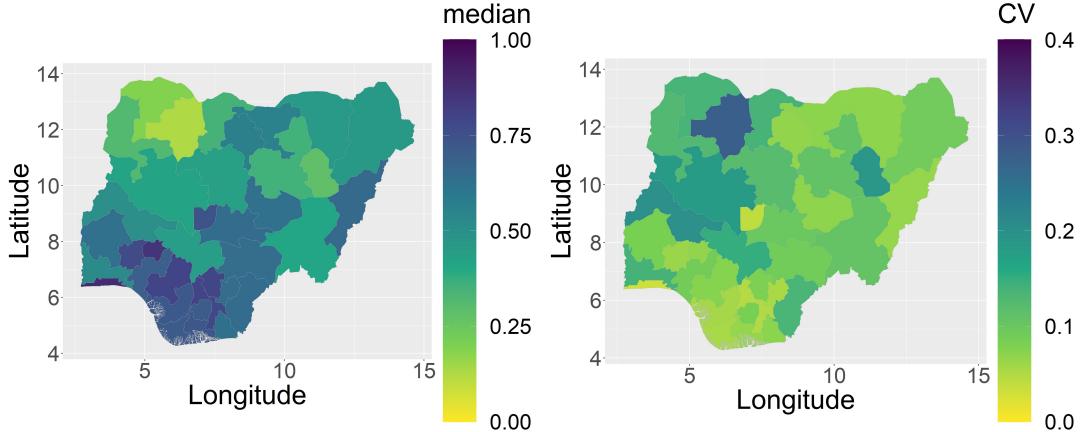


Figure 7: Plot showing the empirically calculated median and coefficient of variation (CV) using our simulated values with a normal prior.

We know that

$$p(\mathbf{X}, \mathbf{Y}) \propto \exp\left(\frac{1}{2} [x^T \ y^T] \begin{bmatrix} \mathbf{Q}_{xx} & \mathbf{Q}_{xy} \\ \mathbf{Q}_{yx} & \mathbf{Q}_{yy} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}\right).$$

Hence, by comparison we see that

$$\mathbf{Q}_{xx} = \tau \mathbf{R} + \mathbf{Q}, \quad \mathbf{Q}_{xy} = \mathbf{Q}_{yx}^T = -\mathbf{Q}, \quad \mathbf{Q}_{yy} = \mathbf{Q}.$$

Then it follows from known results that $p(\mathbf{X}|\mathbf{Y})$ is normally distributed with the following expectation and precision matrix:

$$\mu_{\mathbf{X}|\mathbf{Y}} = -\mathbf{Q}_{xx}^{-1} \mathbf{Q}_{xy} \mathbf{Y} = [\tau \mathbf{R} + \mathbf{Q}]^{-1} \mathbf{Q} \mathbf{Y}, \quad \mathbf{Q}_{\mathbf{X}|\mathbf{Y}} = \mathbf{Q}_{xx} = \tau \mathbf{R} + \mathbf{Q}$$

This is a proper GMRF, since the precision matrix has rank 37. Intuitively, we understand it as intrinsic GMRFs model intrinsic behaviour. That is, they model intrinsic fluctuations around some unknown mean. By conditioning on data, as done above, we add information about this mean, making it a proper GMRF, and not only an intrinsic GMRF.

As we can see from Figure 7 and Figure 8, the model with the Besag prior has a lower span in the median compared to the model with the vague prior. We can also see in Figure 8 that the coefficients of variation are generally lower for the model with the Besag prior. This indicates that we were successful in reducing the variance by introducing bias, as expected.

2.4 d)

We now want to update our observations, using an independent survey from *Kaduna* with low variance. We update our vector $\tilde{\mathbf{Y}} = (Y_1, \dots, Y_{37}, Y_{Kaduna})$. Before we assumed $\mathbf{Y} = \mathbf{X} + \varepsilon$, but

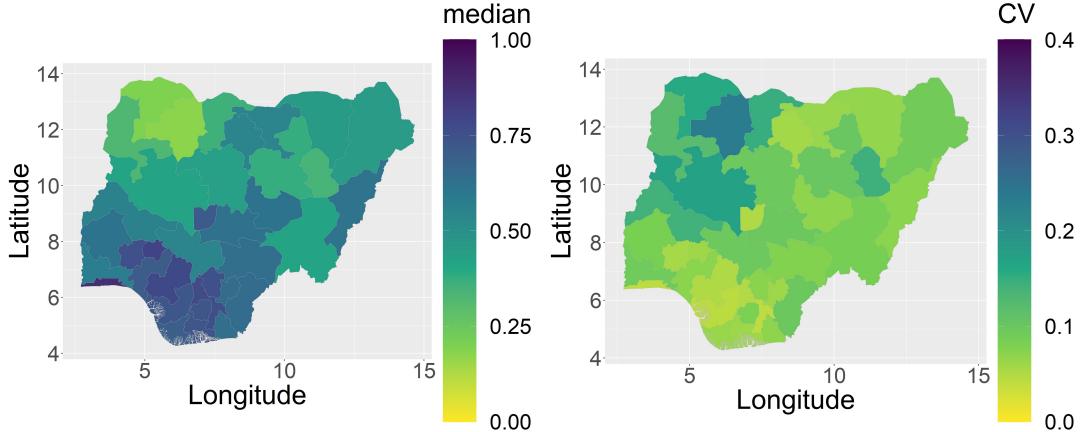


Figure 8: Plot showing the empirically calculated median and coefficient of variation (CV) using our simulated values with a Besag prior.

now we also want to model the *Kaduna* region through conditioning on the independent survey. Therefore, we will instead assume $\tilde{\mathbf{Y}} = A\mathbf{X} + \tilde{\varepsilon}$, where

$$A = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 1 \\ 0 & \dots & 1 & \dots & 0 \end{bmatrix},$$

and the 1 in the last row is in column 17, indicating the Kaduna region. The $\tilde{\varepsilon}$ we use in this case is a modified version of the previous one, consisting V_i 's. We now include a last component, namely 0.1^2 , coming from the precise, independent survey in Kaduna.

The distribution of $\tilde{\mathbf{Y}}|\mathbf{X}$ is given by a normal distribution with mean $A\mathbf{X}$ and covariance matrix given by $\tilde{\Sigma}$, where $\tilde{\Sigma}$ is a diagonal matrix that has the V_i 's for $i = 1, \dots, 37$ followed by 0.1^2 on the diagonal. That is,

$$\tilde{\Sigma} = \begin{bmatrix} V_1 & 0 & \dots & 0 & 0 \\ 0 & V_2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & V_{37} & 0 \\ 0 & 0 & \dots & 0 & 0.1^2 \end{bmatrix}.$$

We denote its precision matrix by $\tilde{Q} = \tilde{\Sigma}^{-1}$.

To find the distribution of $\mathbf{X}|\tilde{\mathbf{Y}}$, we take a similar approach as we did in section 2.3. As before, we first compute $p(\mathbf{X}, \tilde{\mathbf{Y}}) = p(\mathbf{X})p(\tilde{\mathbf{Y}}|\mathbf{X})$.

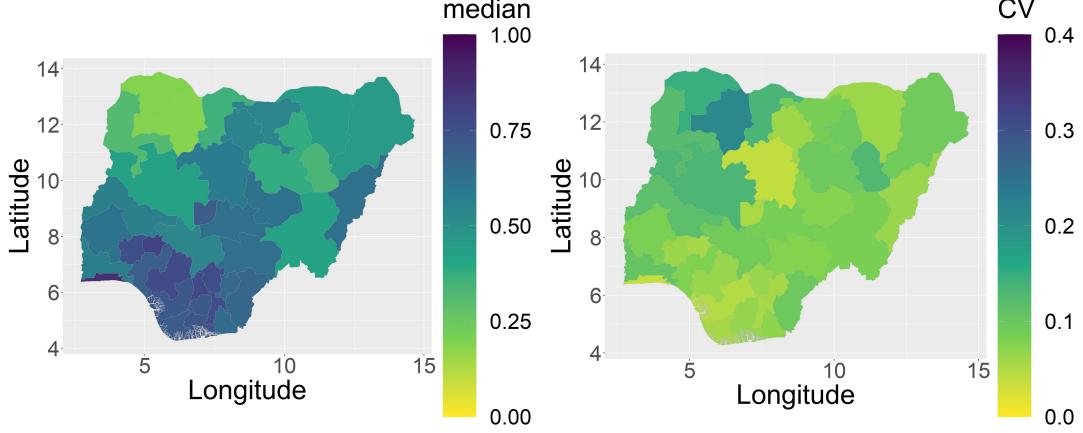


Figure 9: Plot showing the empirically calculated median and coefficient of variation (CV) using our simulated values with a Besag prior and with an extra, independent survey for the Kaduna region.

$$\begin{aligned}
p(\mathbf{X}, \tilde{\mathbf{Y}}) &\propto \exp\left(-\frac{1}{2}(\tilde{\mathbf{y}} - A\mathbf{x})^T \tilde{\mathbf{Q}}(\tilde{\mathbf{y}} - A\mathbf{x})\right) \times \exp\left(-\frac{1}{2}\mathbf{x}^T \tau \mathbf{R} \mathbf{x}\right) \\
&= \exp\left(-\frac{1}{2}\mathbf{x}^T \tau \mathbf{R} \mathbf{x} - \frac{1}{2}\tilde{\mathbf{y}}^T \tilde{\mathbf{Q}} \tilde{\mathbf{y}} + \frac{1}{2}(A\mathbf{x})^T \tilde{\mathbf{Q}} \tilde{\mathbf{y}} + \frac{1}{2}\tilde{\mathbf{y}}^T \tilde{\mathbf{Q}} A\mathbf{x} - \frac{1}{2}(A\mathbf{x})^T \tilde{\mathbf{Q}} A\mathbf{x}\right) \\
&= \exp\left(-\frac{1}{2}\mathbf{x}^T (\tau \mathbf{R} + A^T \tilde{\mathbf{Q}} A)\mathbf{x} + \frac{1}{2}(A\mathbf{x})^T \tilde{\mathbf{Q}} \tilde{\mathbf{y}} + \frac{1}{2}\tilde{\mathbf{y}}^T \tilde{\mathbf{Q}} A\mathbf{x} - \frac{1}{2}\tilde{\mathbf{y}}^T \tilde{\mathbf{Q}} \tilde{\mathbf{y}}\right)
\end{aligned}$$

Hence, by a comparison similar to the one in section 2.3 we see that

$$\tilde{\mathbf{Q}}_{xx} = \tau \mathbf{R} + A^T \tilde{\mathbf{Q}} A, \quad \tilde{\mathbf{Q}}_{xy} = \tilde{\mathbf{Q}}_{yx}^T = -A^T \tilde{\mathbf{Q}}, \quad \tilde{\mathbf{Q}}_{yy} = \tilde{\mathbf{Q}}.$$

Then it follows from known results that $p(\mathbf{X}|\tilde{\mathbf{Y}})$ is normally distributed with the following expectation and precision matrix:

$$\mu_{\mathbf{X}|\tilde{\mathbf{Y}}} = -\tilde{\mathbf{Q}}_{xx}^{-1} \tilde{\mathbf{Q}}_{xy} \tilde{\mathbf{Y}} = [\tau \mathbf{R} + A^T \tilde{\mathbf{Q}} A]^{-1} A^T \tilde{\mathbf{Q}} \tilde{\mathbf{Y}}, \quad \tilde{\mathbf{Q}}_{\mathbf{X}|\tilde{\mathbf{Y}}} = \tilde{\mathbf{Q}}_{xx} = \tau \mathbf{R} + A^T \tilde{\mathbf{Q}} A.$$

As we observe in Figure 9, where we again simulate 100 realizations and compute the empirical medians and coefficients of variation, the Kaduna region clearly has a lower coefficient of variation than from the previous model shown in Figure 8. Furthermore, the neighboring states are also influenced through the spatial structure of the model, yielding lower variance for these regions as well. If we were to do a similar survey and update the model from section 2.2, we would not see any decrease in the neighboring states, since we assumed them to be independent in that prior. Therefore, the gain is higher when adding additional information to spatial models, as the information added to one region admits a flow of variance reduction through the neighbors.

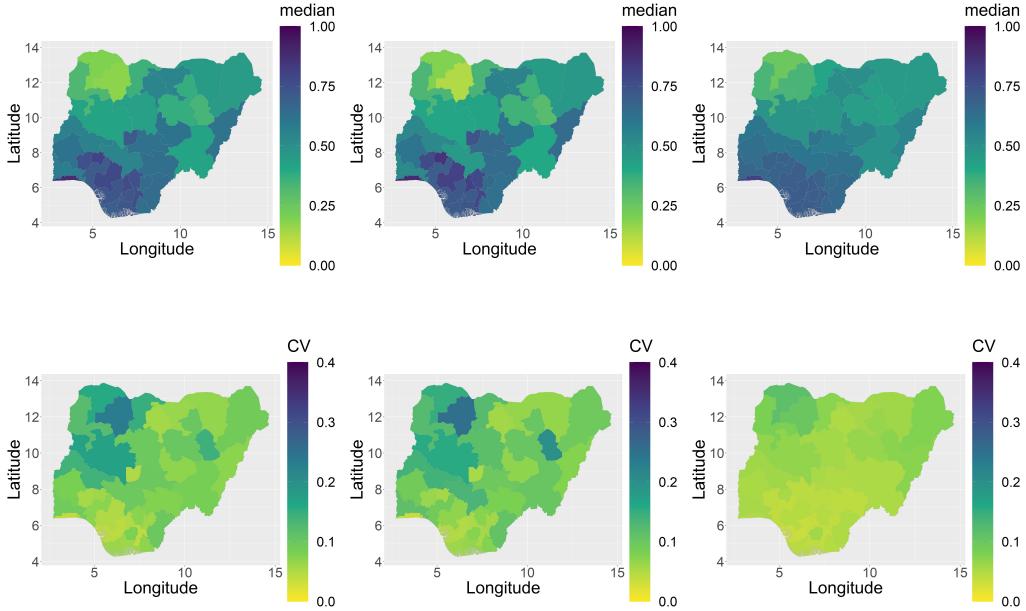


Figure 10: Plot showing the empirically calculated median and coefficient of variation (CV) using three different values of τ : 1 (left), 0.1 (center), and 10 (right).

2.5 e)

In this problem, we investigate the effect of changing the precision parameter, τ .

As we can see in Figure 10, a higher τ yields a lower coefficient of variation. Since the precision parameter τ regulates the impact of the structure matrix \mathbf{R} , which we know is given from the graph structure, we suspect that a higher τ increases the bias even more to reduce variation. One can of course ask whether or not it is reasonable to increase bias even further to obtain lower variance. We consider it to be unreasonable to use $\tau = 10$ for example, as the observed proportions shown in Figure 6 does not look similar to the median plot in Figure 10 with $\tau = 10$. In addition, as we get a steadily low coefficient of variation, it seems unlikely that we will get as much fluctuations between the regions as we saw in Figure 6. On the other hand, we see that both the median plots for $\tau = 0.1$ and $\tau = 1$ seem more reasonable, where the latter has more bias, and therefore less variation. For $\tau = 0.1$ we see that our results are similar to what we got in Figure 7. This makes sense, as both the normal prior (the vague prior) we used in problem 2.2 and a Besag model with low τ -value introduces less bias, and therefore the observed results will have a more prominent impact on the final estimates.

If we had to chose some value for τ based on this, we'd pick a value for τ somewhere in the interval $[0.1, 1]$. Intuitively, we consider it to be reasonable to have estimated a higher τ if we were to use the admin2-graph, as the admin1-graph is too coarse to suggest such strong dependence between regions.

Looking at the plots in Figure 10 comparing $\tau = 10$ to $\tau = 0.1$ and $\tau = 1$, we consider it important to estimate τ correctly, at least to a precise order of magnitude. By what we see in Figure 10,

it looks like the precision parameter τ controls the bias-variance trade-off. Therefore, if we fail to estimate τ precisely, we will probably experience too much bias, trying to introduce a model depending too much on the spatial facts, or too much variance, having a flexible model, but with a lot of variation.

If we were to chose $\tau = 0$, we observe that we get a model with independent regions.

2.6 f)

The log likelihood $l(\tau; \mathbf{y}) = \log(f(\mathbf{y}; \tau))$, and from Bayes' theorem we know that

$$f(\mathbf{x}|\mathbf{y}) = \frac{f(\mathbf{x})f(\mathbf{y}|\mathbf{x})}{f(\mathbf{y})} \implies f(\mathbf{y}) = \frac{f(\mathbf{x})f(\mathbf{y}|\mathbf{x})}{f(\mathbf{x}|\mathbf{y})}.$$

Hence,

$$\begin{aligned} \log(f(\mathbf{y}; \tau)) &= \log\left(\frac{f(\mathbf{x}; \tau)f(\mathbf{y}|\mathbf{x})}{f(\mathbf{x}|\mathbf{y}; \tau)}\right) \\ &= \log(f(\mathbf{x}; \tau)) + \log(f(\mathbf{y}|\mathbf{x})) - \log(f(\mathbf{x}|\mathbf{y}; \tau)) \end{aligned}$$

It is not possible to directly compute the marginal $f(\mathbf{y}; \tau)$ by solving $\int f(\mathbf{x}, \mathbf{y}; \tau) d\mathbf{x}$, since $f(\mathbf{x}, \mathbf{y}; \tau)$ is a improper prior, and therefore our integral would be infinite.

To compute $l(\tau; \mathbf{y})$, we consider the formula above where we insert the expressions given. As we know, \mathbf{X} is a Besag model with precision matrix $\tau\mathbf{R}$ and $\mathbf{Y}|\mathbf{X}$ is normally distributed with mean \mathbf{X} and covariance matrix \mathbf{D} , which we previously denoted by Σ . We also know that $\mathbf{X}|\mathbf{Y}$ is normally distributed with mean $\mu_C = [\tau\mathbf{R} + \mathbf{Q}]^{-1}\mathbf{Q}\mathbf{Y}$ and with precision matrix $\mathbf{Q}_C = \tau\mathbf{R} + \mathbf{Q}$ as in section 2.3.

We therefore obtain

$$\begin{aligned} \log(f(\mathbf{y}; \tau)) &= \log(f(\mathbf{x}; \tau)) + \log(f(\mathbf{y}|\mathbf{x})) - \log(f(\mathbf{x}|\mathbf{y}; \tau)) \\ &= \text{Const} + \log\left(\tau^{(37-1)/2} \exp\left(-\frac{1}{2}\mathbf{x}^T\tau\mathbf{R}\mathbf{x}\right)\right) \\ &\quad + \log\left(\exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{x})^T\mathbf{D}^{-1}(\mathbf{y} - \mathbf{x})\right)\right) \\ &\quad - \log\left(|\mathbf{Q}_C^{-1}|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{y} - \mu_C)^T\mathbf{Q}_C(\mathbf{y} - \mu_C)\right)\right) \\ &= \text{Const} + 18\log(\tau) - \frac{\tau}{2}\mathbf{x}^T\mathbf{R}\mathbf{x} - \frac{1}{2}(\mathbf{y} - \mathbf{x})^T\mathbf{D}^{-1}(\mathbf{y} - \mathbf{x}) \\ &\quad + \frac{1}{2}\log(|\mathbf{Q}_C^{-1}|) + \frac{1}{2}(\mathbf{x} - \mu_C)^T\mathbf{Q}_C(\mathbf{x} - \mu_C), \end{aligned}$$

which is exactly what we wanted.

When maximising the likelihood in R, we obtain the numerical answer $\hat{\tau} = \tau_{MLE} = 0.8062853$ as the MLE for τ . This seems fair, as we discussed in section 2.5 that both $\tau = 0.1$ and $\tau = 1$ would seem somewhat reasonable. The likelihood is equal for all x , and hence we just chose an x in our computation.

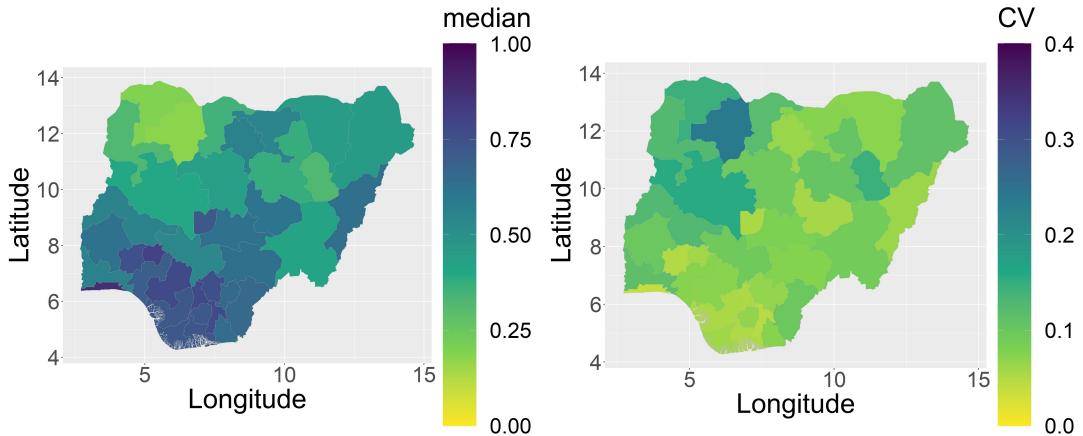


Figure 11: Plot showing the empirically calculated median and coefficient of variation (CV) using our simulated values with a Besag prior and precision parameter $\tau = \hat{\tau} = \tau_{MLE}$.

Compared to what we saw in Figure 10, the results where we use $\tau = \hat{\tau}$, as shown in Figure 11, look quite promising. In Figure 10, it may look like the model with $\tau = 0.1$ could benefit from introducing some more bias to reduce variance, while the model with $\tau = 1$ might be a bit too biased. The plot of the median in Figure 11 indicates that we probably have found a sweet spot in the bias-variance trade-off, where the coefficient of variation indicates less fluctuations than for $\tau = 0.1$ without introducing too much bias, and less bias than for $\tau = 1$ without introducing too much variance.